Title Page: Association of microRNA 17-92 cluster host gene (MIR17HG) polymorphisms with breast cancer

Running Title: Association of MIR17HG SNPs with breast cancer

Authors and affiliations

Diego Chacon-Cortes[1, 2], Robert A Smith[1, 2], Rodney A. Lea[1, 2], Philippa H Youl[2, 3,4] and Lyn R. Griffiths[1, 2*].

[1]Genomics Research Centre, Institute of Health and Biomedical Innovation, Queensland University of Technology, Musk Avenue, Kelvin Grove. Queensland 4059, Australia.

[2]Griffith Health Institute, Griffith University, Queensland, Australia

[3]Cancer Council Queensland, Brisbane, Queensland, Australia

[4]School of Public Health, Queensland University of Technology, Queensland, Australia

*Address for correspondence:

Professor Lyn Griffiths

Genomics Research Centre

Institute of Health and Biomedical Innovation

Queensland University of Technology

60 Musk Avenue, Kelvin Grove 4059

Queensland, Australia.

lyn.griffiths@qut.edu.au

Phone: + 61 7313 86102

Fax: + 61 7313 6039

Abstract

Breast cancer incidence and mortality rates are increasing despite our current knowledge on the disease. 95% of breast cancer cases correspond to sporadic forms of the disease and are believed to involve an interaction between environmental and genetic determinants. The microRNA 17-92 cluster host gene (MIR17HG) has been shown to regulate expression of genes involved in breast cancer development and progression. Study of single nucleotide polymorphisms (SNPs) located in this cluster gene could help provide a further understanding of its role in breast cancer. Therefore, this study investigated 6 SNPs in the MIR17HG using two independent Australian Caucasian case-control populations (GRC-BC and GU-CCQ BB populations) to determine association to breast cancer susceptibility. Genotyping was undertaken using chip-based matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS). We found significant association between rs4824505 and breast cancer at the allelic level in both study cohorts (GRC-BC $p = 0.01$ and GU-CCQ BB $p = 0.03$). Furthermore haplotypic analysis of results from our combined population determined a significant association between rs4824505/rs7336610 and breast cancer susceptibility ($p = 5x10^{-4}$). Our study is the first to show that the A allele of rs4824505 and the AC haplotype of rs4824505/rs7336610 are associated with risk of breast cancer development. However, definitive validation of this finding requires larger cohorts or populations in different ethnical backgrounds. Finally, functional studies of these SNPs could provide a deeper understanding of the role that MIR17HG plays in the pathophysiology of breast cancer.

Keywords: association analysis / breast cancer / haplotype / microRNA / single nucleotide polymorphisms (SNP)

Introduction

In 2012, 14.09 million people in the world were newly diagnosed with cancer and about 21% of them corresponded to breast cancer (1 676 633 cases in total) according to the International Agency for Research on Cancer (IARC). It is the most common type of cancer in women worldwide and is estimated to be developed by one in eight women living to the age of 90. Breast cancer is also the fifth cause of cancer-related deaths in the world (6.4%) and accounted for 197 528 of all cancer related deaths in developed countries. It has become a cause for concern since both incidence and mortality rates have increased by more than 20% and 14% respectively from figures reported in 2008[1-3]. Sporadic forms of breast cancer account for about 95% of all cases and they are most likely the result of interaction between genetic and environmental factors [4-6, 3].

MicroRNAs (miRNAs) are small non-coding single stranded RNA molecules of about 21 to 25 nucleotides in length involved in gene regulation at the translational level. They account to about 1% of the human genome and are very likely to be important in cancer biology due to their role in regulating important cellular processes like cell growth, differentiation and cell survival [7-9, 3, 10]. miRNAs are usually generated following the canonical pathway where they are transcribed from long microRNA primary transcripts (pri-miRNAs) about 1kb in size that contain up to six miRNA precursors (pre-miRNA). Pre-miRNAs are stem-loop molecules of about 55 to 70 nucleotides in length and they are produced when pri-miRNA are cleaved by the complex Drosha – DCGR8 (DiGeorge syndrome critical region gene 8, also known as pasha) within the cellular nucleus. Pre-miRNAs are then transported to the cytoplasm by XPO5 (exportin5) and the nuclear protein Ran-GTP. In the cytoplasm pre-miRNAs are processed by DICER1, a ribonuclease type III enzyme, and double-stranded RNA binding proteins (RBP) TARBP2 and/or PKRA producing a duplex molecule that contains both the single-stranded mature miRNA sequence and its complementary strand named miRNA*. The miRNA:miRNA* molecule is loaded into the RNA-induced silencing complex (RISC), where the mature miRNA sequence is merged into the Argonaute/EIFC2C (Ago) proteins and the miRNA* is released and degraded. Ago proteins bound to mature miRNA mediate target messenger RNA (mRNA) recognition and interaction between these three components results in gene regulation. The mRNA target is recognised by pairing of the miRNA seed region (nucleotides 2 to 8) located in the 5'-end of the miRNA and the complementary sequence mainly located in the 3'-UTR region of the mRNA[11-14]. Each miRNA may bind to up to 200 gene targets and each gene could have multiple binding sites for different miRNAs [15, 16]. Location of promoter regions of microRNA genes, mechanisms that regulate their biogenesis and molecular mediators of their functional effects within cells and tissues are still being studied and remain unclear for most miRNAs identified to date [17, 18].

Around 52% of all miRNA genes are located in different non-random positions within the genome, generally in regions of chromosomal instability or cancer-associated regions [19]. Changes in miRNA synthesis, expression and effect on target genes in relation to cancer could occur through many different mechanisms and some of them include: point mutations in miRNA genes, mRNA sequences and surrounding regions, epigenetic changes such as mRNA gene methylation, loss or mutation in the promoter regions for specific miRNA clusters and/or alterations in pathway related RBP [8, 17].

Approximately one-third of all miRNAs form genomic clusters, about 113 in total, which may measure up to 51 kb in length, and they are usually transcribed as a single polycistronic transcript [10, 8]. One such cluster is the

microRNA 17-92 cluster host gene (MIR17HG) located on chromosome 13q31.3 in the third intron of the *c13orf25* gene. It was initially identified in relation to B cell lymphoma both in cell lines and tissue sample from patients by Ota et al in 2004[20]. The MIR17HG cluster transcript is about 800bp in length and contains six miRNAs: MIR17, MIR18A, MIR19A, MIR20A, MIR19B1 and MIR92A1, which belong to 4 seed families (miRNA 17, miRNA 18, miRNA 19 and miRNA 92). Based on their seed sequences, two related paralogues of this cluster have also been identified in the human genome: microRNA 106b 25 cluster located on chromosome 7 (7q22.1) in the 13th intron of gene MCM7 and microRNA 106a-363 cluster located on chromosome X (Xq26.2) [21].

The MIR17HG was named oncomiR-1 by He et al who found it to promote oncogenesis in B-cell lymphoma using a murine model [22]. Further studies have confirmed roles both in cell proliferation and cell death, a feature that seems to be dependent on the cell types and tissues in which it is found. Research has shown that the MIR17HG inhibits tumour growth by targeting important cell cycle regulator genes including E2F1, MYC, RB1 and CCND1. In 2005, O'Donnell et al observed that MIR17HG down-regulated expression of both E2F1 and MYC genes using Burkitt lymphoma cells (P493-6) as a model [23]. On the other hand, later studies have shown this cluster gene region seems to have an oncogenic effect in other types of tumour. Experimental models of human B-cell lymphomas overexpressing MIR17HG showed evasion of apoptosis due to enhanced activity of MYC. Moreover, studies of small-cell lung carcinomas showed negative regulation of the pro-apoptotic gene BCL2L11 resulting from overexpression of MIR17HG [24]. This finding may suggest that fine regulation of MIR17HG expression is important. Research studies indicate that specific processing of miRNAs included in this cluster might be the result of largely unknown intricate regulatory mechanisms. There are 34 transcription factors that regulate MIR17HG transcription identified to date and they include MYC, MYCN, MXI, E2F3 and TP53 amongst others, that are specific to different cell types and biological contexts [25, 26].

In relation to breast malignancies, a number of studies on the effects of MIR17HG in breast cancer cell lines have also been published. In 2006 MIR17 5p, one of the miRNAs included in the cluster, was shown to inhibit oestrogen receptor α (ESR1) co-activator NCOA3 in different breast cancer cell lines [27]. Following this finding, Yu et al demonstrated a tumour suppressor role for MIR17HG inducing cell cycle arrest and decreasing cell proliferation by directly inhibiting CCND1 in the MCF-7 breast cancer cell line in 2008[28]. Similarly in a later study, Leivonen et al showed that MIR17HG reduced cell growth and progression of cell cycle in breast cancer cell lines MCF-7 and BT-474 using two mechanisms: translational repression of ESR1 and/or

downregulation of ESR1 responding genes. Additionally, they were able to confirm their *in-vitro* findings through expression studies on breast cancer tumour samples showing significant association between high expression levels of MIR18a and ESR1-negative breast cancer tumours[29].

According to the evidence previously presented, the MIR17HG seems to play an important role in development of different carcinomas including breast cancer. However, there is a lack of knowledge on the molecular mechanisms leading to altered expression or regulation of this particular cluster as well as each of the mature miRNAs included in it. Genetic variants such as single nucleotide polymorphisms (SNPs) could potentially impact biological processes involved in the production or functional effects of the microRNA 17-92 cluster host gene and the study of such variants in relation to cancer could provide insight into its aetiological mechanisms.

In this study, we genotyped 6 SNPs located in the cluster gene region or in close proximity to it. Genotyping of these variants was performed using multiplex PCR and Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis in two independent Australian Caucasian breast cancer cohorts. Results from our allele and haplotype association analysis showed statistical significance for two of the selected SNPs and the risk of breast cancer in these Caucasian populations.

Materials and methods

Study Populations

Genotyping was carried out in two independent cohorts of females of Caucasian (Northern European) origin: The Genomics Research Centre Breast Cancer (GRC-BC) population and a subset of the Griffith University-Cancer Council Queensland Breast Cancer Biobank (GU-CCQ BB). GRC-BC population was recruited from the Gold Coast Hospital, Southport and it included 244 breast cancer patient samples from patients residing in the South East Queensland region. Samples from 187 healthy females with no history of personal or familial cancer were used as controls and they were also obtained via the Genomics Research Centre Clinic, Southport, with the research approved by Griffith University's Human Ethics Committee (Approval: MSC/07/08/HREC and PSY/01/11/HREC) and the Queensland University of Technology Human Research Ethics Committee (Approval: 1400000104). Mean age of the test populations was 57.52 years and 57 years, for cases and controls, respectively.

Replication was performed using 679 samples from the GU-CCQ BB population. Patient samples were collected by the Genomics Research Centre in collaboration with the Cancer Council of Queensland as part of a 5-year population-based longitudinal study of women newly diagnosed with breast cancer. 929 women resident in Queensland with a diagnosis of invasive breast cancer confirmed histologically have been recruited by this biobank since January 2010. Age range of patient samples included in this study varied from 33 to 80 years, with an average age of 60.16. Clinical and demographic information was obtained through the Queensland Cancer Registry, Genomics Research Centre General Questionnaire, computer assisted telephone interview (CATI) and self-administered questionnaires (SAQ) over two main time points: 4 to 6 months post diagnosis and 18 months post diagnosis [approximately 12 months after initial interview]. Additionally, using the Queensland Cancer Registry and the National Death Index, the study cohort will be followed-up for five years post diagnosis to record cancer recurrence and mortality [30]. 308 ethnicity and sex matched control samples for the GU-CCQ BB population were obtained from 2 sources: 107 females with no personal or familial history of cancer, recruited from January 2000 through the Genomics Research Centre at the Institute of Health and Biomedical Innovation Institute, Queensland University of Technology and also genotyping data obtained from 201 individuals belonging to the phase 1 European population from the 1000Genomes project [31].

Genomic DNA sample preparation from whole human blood.

Genomic DNA was obtained from whole blood samples using a modified salting out method [32, 33]. DNA samples were checked for quantity and purity using spectrophotometry measurements using the Thermo Scientific NanoDrop^{TM} 8000 UV-Vis Spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE. USA) to determine concentration and 260/280 ratios [34-36]. To guarantee DNA sample quality, those with a value below 1.7 for their 260/280 ratio were further purified using an ethanol precipitation protocol [37].

miRNA SNP selection

We used the human miRNA disease database (HMDD), developed by Lu et al [38] and updated in January 2012, to identify microRNAs involved in development and progression of breast cancer. We selected 24 biological/cellular functions and 8 diseases and/or pathological features related to breast cancer from two datasets included in the HMDD: “The whole miRNA-disease association data” and “The miRNA function set data”. The microRNA 17-92 cluster host gene (MIR17HG) was found to be present in most of the features on each dataset and therefore we conducted a preliminary literature search to determine its role in cancer

development and progression and to find SNP genotyping studies for this cluster gene. Following this, we searched for previously identified SNPs in the MIR17HG using both the dbSNP database from The National Center for Biotechnology Information (NCBI) [39] and 1000 Genomes project browser [31]. MicroRNA SNPs (miR-SNPs) that were located inside the pre-miRNA gene were automatically included in a preliminary selection and those located outside of the pre-miRNA gene were selected using either one of the following criteria: SNPs located within 500bp to the 3' or 5' end of the pre-miRNA gene region were automatically included in the preliminary selection. Finally, miR-SNPs located at a distance higher than 500bp from either end of the pre-miRNA gene were chosen only if they had a minor allele frequency (MAF) higher than 0.05 in Caucasian populations. Our preliminary selection identified 13 miR-SNPs in the MIR17HG or located up to 1.6 kb downstream and two variants (rs72631821 and rs9589207) located inside one of the mature miRNA sequences (MIR92A1) (See Table 1).

Primer design and multiplex PCR genotyping

We were able to design a multiplex PCR assay to include six miR-SNPs: one of the variants located inside the MIR92A1 gene (rs9589207) and five miR-SNPs located throughout the MIR17HG region using the MassARRAY® Assay Design Suite v1.0 software (SEQUENOM Inc., San Diego, CA, USA). Two PCR primers (forward and reverse) and one iPLEX® (extension) primer were designed for each miR-SNP to achieve successful marker and allele identification by mass spectrometry (See Table 2). We verified that masses of extension primers differed by at least 30 Da among different SNPs and by 5 Da between alternative alleles of the same marker and primers were synthesized by Integrated DNA Technologies (IDT®) Pte. Ltd. (Baulkham Hills, NSW 2153, Australia). Multiplex PCR reactions were performed according to the manufacturer's protocol for the iPLEX™ GOLD genotyping application using iPLEX® GOLD reaction kit reagents (SEQUENOM Inc., San Diego, CA, USA). Primer extension reactions were carried out using the SEQUENOM linear adjustment method as per manufacturer's iPLEX® GOLD genotyping protocol. All iPLEX® GOLD genotyping reactions were performed using Applied Biosystems® MicroAmp® EnduraPlate™ Optical 96-Well Clear Reaction Plates with Barcode (Life Technologies Australia Pty Ltd., Mulgrave, VIC, Australia) and a Applied Biosystems® Veriti® 96-Well Thermal Cycler (Life Technologies Australia Pty Ltd., Mulgrave, VIC, Australia).

MALDI-TOF MS analysis and data analysis

We dispensed a total of 12-16 nl of each iPLEX® reaction product onto a SpectroCHIP® II G96 (SEQUENOM Inc., San Diego, CA, USA) using SEQUENOM® MassARRAY® Nanodispenser (SEQUENOM Inc., San Diego, CA, USA) and SpectroCHIP® analysis was performed by SEQUENOM® MassArray® Analyzer 4 (SEQUENOM Inc., San Diego, CA, USA). We acquired automated spectra after laser desorption/ionisation using the SpectroAcquire software version 4.0 (SEQUENOM Inc., San Diego, CA, USA) and genotype data analysis was performed using the MassARRAY® Typer software version 4.0 (SEQUENOM Inc., San Diego, CA, USA).

Statistical Analysis

We determined genotype and allele frequencies for each miRNA SNP in our case and control populations. Hardy-Weinberg equilibrium (HWE)[40, 41] was used to evaluate deviation between observed and ideal Hardy-Weinberg frequencies. Differences in genotype and allele frequencies between cases and controls were evaluated using Chi-square analysis[42] for each independent population to determine association with breast cancer. The α for p-values was set at 0.05 to determine statistical significance. Statistical analysis of genotypes and alleles was conducted using Plink software version 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) [43]. We also calculated odds ratio (OR) and a confidence interval (CI) of 95% to assess disease risk. Finally, we merged results from both populations and we performed linkage disequilibrium (LD) and haplotype block association analysis on the combined genotyping data using Haploview software version 4.2 (Daly Lab, Broad Institute, Cambridge, MA, USA) [44].

Results

Genotypic Analysis

Table 3 shows the 6 SNPs genotyped in our study provided good coverage of the MIR17HG. They also included rs9589207 located inside a mature miRNA sequence (MIR92A1) and rs1888138 located 578 bp downstream of the cluster gene. All SNPs had a MAF greater than 0.05, with the exception of rs9589207 which had a MAF of 0.02. We proceeded to include all these SNPs in our multiplex PCR, in particular rs9589207 because of their genetic location. However, rs9589207 was finally excluded from our association analysis because we were unable to find the mutant allele in our initial genotyping of the GRC-BC cohort.

Hence genotyping results of the 5 remaining SNPs obtained from both populations are shown in Table 4. All control populations were in HWE ($p > 0.05$) for all tested SNPs and allele and genotype frequencies closely matched those found in Hapmap for Caucasian populations. As shown in Table 4, we were able to validate significant difference in allele frequencies between cases and controls for rs4284505 in both the GRC-BC and GU-CCQ BB populations using chi-square analysis ($p = 0.01$ and 0.03 respectively). Statistical analysis of allele frequencies in cases and controls in the GRC-BC population for rs7336610 showed significance ($p = 0.03$), but we were unable to obtain a similar finding in our independent replication population ($p = 0.10$) and all the other remaining SNPs showed no significant differences between cases and controls in both populations. We observed lower frequencies (38.3% and 35.9% for GRC-BC and GU-CCQ cohorts respectively) of the A allele for rs4284505 in cases for both independent populations (See Table 4) and we calculated an odds ratio of 0.70 (CI 95%: 0.52 – 0.93) for GRC-BC population and of 0.79 (CI 95%: 0.65 – 0.97) for GU-CCQ population. These results indicate that the presence of the A allele in this loci seems to have a protective effect on susceptibility to breast cancer.

Finally, chi square analysis of genotypes for all selected MIR17HG SNPs between cases and controls showed significant differences for rs4284505 and rs7336610 ($p = 0.01$ for both SNPs) in the GRC-BC population. Interestingly, we were only able to obtain statistical significance in the analysis at the genotype level only for rs7336610 in our replication cohort ($p = 0.04$).

Haplotypic Analysis

Following this, we combined genotyping results from both GRC-BC and GU-CCQ populations to conduct linkage disequilibrium (LD) and haplotype block association analysis. As shown in Figure 1, two of our selected SNPs (rs4284505 and rs7336610) were part of a 3kb LD block (Block 1) with a D' score of 0.98. Table 5 shows frequencies for the different haplotypes in block 1 found in cases and controls, as well as results of haplotype association analysis to breast cancer using chi-square analysis. According to our findings, cases had a lower frequency of haplotype AC (36.4%) than controls (43.3%) and statistical analysis proved this haplotype to be the most significantly associated with breast cancer ($p = 5x10^{-4}$). Finally, we calculated an odds ratio for haplotype AC of 0.75 with a 95% confidence interval of 0.60 to 0.94 ($p = 0.012$). Results from our haplotype analysis validate our finding for the genotyping analysis at the allele level in rs7336610. These results also suggest that for the individuals that carry the AC haplotype, it appears to confer a protective effect on risk of

breast cancer in Caucasians reducing the risk of being affected breast cancer to about 25% in comparison to other haplotype carriers.

Discussion

Breast cancer affects a significant number of women and mortality rates seem to be increasing, despite substantial research and our current knowledge of the disease [1, 2]. MicroRNAs are involved in molecular pathways leading to cell growth, differentiation and survival and there is enough evidence to show they play a significant role in the mechanisms that lead to development and progression of different types of cancer. Changes in expression of the microRNA 17-92 cluster host gene, also known as oncomiR-1, were initially described in haematopoietic and lung cancers where it was shown to play a role in tumour growth and apoptosis[20-24]. Research has also highlighted MIR17HG as an important regulator of genes involved in breast cancer pathways. Studies by Hossain et al and Leivonen et al showed it plays a role in regulation of oestrogen receptor α (ESR1) in different breast cancer cell lines and tumour tissues through transcriptional downregulation of this receptor and genes that interact or co-activate ESR1 [27, 29]. Yu et al found the MIR17HG to have a tumour suppressor role through CCND1 inhibition in the MCF-7 cell line [28]. However, we currently lack understanding of the specific features that result in altered biogenesis and functional effects of this particular miRNA cluster gene in breast malignancies. Therefore the study of single nucleotide polymorphisms located inside the MIR17HG using well defined breast cancer cohorts may help to assess whether they have a role on the molecular events leading to the development of this disease.

Selection of 6 miRNA SNPs found 0.6 kb downstream and inside the microRNA 17-92 cluster host gene region provided a thorough coverage of the gene and we were able to successfully genotype all of these variants in the two available cohorts, even rs9589207 located inside the MIR92A1 transcript. Unfortunately this particular SNP was not included in our association analysis because we could not identify the presence of the mutant allele in any of our populations. However two of the five remaining SNPs, rs4248505 located about 1.4 kb downstream from MIR17A and rs7336610 found about 1.4 kb upstream from MIR92A1, showed some interesting results on our association analysis of breast cancer risk. Moreover, there does not appear to be any previous genotyping studies on these variants in relation to breast cancer or other diseases in Caucasian or any other ethnical populations. Statistical analysis of our results for rs4284505 showed a significantly higher presence of the A allele in healthy individuals for both GRC-BC and GU-CCQ BB populations (p = 0.01 and 0.03 respectively) and a reduction in breast cancer risk of up to 30%. We also found higher frequencies for the C allele of

rs7336610 in controls in our GRC-BC cohort but failed to establish a similar finding in our replication population. Analysis of the haplotypes present in the LD block rs4284505/rs7336610 showed results consistent to our previous finding in the individual SNP analysis, with the presence of the AC haplotype found to have significantly higher frequencies in controls ($p = 5x10^{-4}$) decreasing the risk of developing breast cancer by about 25%.

In conclusion via analysis of SNPs from MIR17HG, we were able to identify significant association between rs4248505 at the allele level and rs4248505/rs7336610 at the haplotype level with susceptibility to breast cancer. To the best of our knowledge this is the first study investigating SNPs in this cluster gene in Caucasian breast cancer populations and the first to identify that presence of the AC haplotype significantly affects the risk of developing breast cancer. However our findings require further validation in larger populations and/or population of different ethnicities, as well as functional studies to determine the role of this haplotype in miRNA expression or molecular pathways leading to the development of breast cancer.

Acknowledgements

This research was funded by research grants from the Griffith Health Institute, the Cancer Council Queensland and funding from the national Health and Medical Research Council. Blood samples from the Griffith University-Cancer Council Queensland Biobank were collected from patients enrolled in Cancer Council Queensland's Breast Cancer Outcomes Study funded by a Cancer Australia grant (#1006339). Dr Youl is supported by an NHMRC Early Career Research Fellowship (#1054038). We thank Professor Suzanne K. Chambers for her assistance in the development of this project. In addition, this study used infrastructure provided by the Australian government EIF Super Science Funds as part of the Therapeutic Innovation Australia – Queensland node project.

References

1. Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. 2013. http://globocan.iarc.fr. Accessed 22/08/2012.

2. Stewart BW, Wild C, International Agency for Research on C, World Health O. World cancer report 2014. vol Book, Whole. Lyon, France: International Agency for Research on Cancer; 2014.

3. Robbins SL, Kumar V, Cotran RS. Robbins and Cotran Pathologic Basis of Disease. Philadelphia, PA: Saunders/Elsevier; 2010.

4. Anderson DE. Familial versus sporadic breast cancer. Cancer. 1992;70(6 Suppl):1740-6.

5. Claus EB, Schildkraut JM, Thompson WD, Risch NJ. The genetic attributable risk of breast and ovarian cancer. Cancer. 1996;77(11):2318-24. doi:10.1002/(sici)1097-0142(19960601)77:11<2318::aid-cncr21>3.0.co;2-z.

6. Martin A-M, Weber BL. Genetic and Hormonal Risk Factors in Breast Cancer. Journal of the National Cancer Institute. 2000;92(14):1126-35. doi:10.1093/jnci/92.14.1126.

7. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006;6(11):857-66. doi:10.1038/nrc1997.

8. Lynam-Lennon N, Maher SG, Reynolds JV. The roles of microRNA in cancer and apoptosis. Biological Reviews. 2009;84(1):55-71. doi:10.1111/j.1469-185X.2008.00061.x.

9. Yu Z, Baserga R, Chen L, Wang C, Lisanti MP, Pestell RG. microRNA, Cell Cycle, and Human Breast Cancer. The American Journal of Pathology. 2010;176(3):1058-64. doi:10.2353/ajpath.2010.090664.

10. Farazi TA, Spitzer JI, Morozov P, Tuschl T. miRNAs in human cancer. The Journal of Pathology. 2011;223(2):102-15.

11. Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: stepwise processing and subcellular localization. The EMBO journal. 2002;21(17):4663-70.

12. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J et al. The nuclear RNase III Drosha initiates microRNA processing. Nature Publishing Group; 2003. p. 415.

13. Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes & development. 2003;17(24):3011-6. doi:10.1101/gad.1158803.

14. Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature. 2005;436(7051):740-4. doi:10.1038/nature03868.

15. Bartel DP. MicroRNAs: Target Recognition and Regulatory Functions. Cell. 2009;136(2):215-33.

16. Rajewsky N. microRNA target predictions in animals. Nat Genet. 2006.

17. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509-24. doi:10.1038/nrm3838.

18. Kim VN. MicroRNA biogenesis: coordinated cropping and dicing. Nature Reviews Molecular Cell Biology. 2005;6(5):376-85. doi:http://dx.doi.org/10.1038/nrm1644.

19. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(9):2999-3004. doi:10.1073/pnas.0307323101.

20. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res. 2004;64(9):3087-95.

21. Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ et al. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. Cell. 2008;132(5):875-86. doi:10.1016/j.cell.2008.02.019.

22. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435(7043):828-33. doi:10.1038/nature03552.

23. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature. 2005;435(7043):839-43. doi:10.1038/nature03677.

24. Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, Tomida S et al. A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. Cancer Res. 2005;65(21):9628-32. doi:10.1158/0008-5472.CAN-05-2352.

25. Mogilyansky E, Rigoutsos I. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. Cell death and differentiation. 2013;20(12):1603-14. doi:10.1038/cdd.2013.125.

26. Olive V, Jiang I, He L. mir-17-92, a cluster of miRNAs in the midst of the cancer network. The international journal of biochemistry & cell biology. 2010;42(8):1348-54. doi:10.1016/j.biocel.2010.03.004.

27. Hossain A, Kuo MT, Saunders GF. Mir-17-5p regulates breast cancer cell proliferation by inhibiting translation of AIB1 mRNA. Molecular and cellular biology. 2006;26(21):8191-201. doi:10.1128/MCB.00242-06.

28. Yu Z, Wang C, Wang M, Li Z, Casimiro MC, Liu M et al. A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation. The Journal of cell biology. 2008;182(3):509-17. doi:10.1083/jcb.200801079.

29. Leivonen SK, Makela R, Ostling P, Kohonen P, Haapa-Paananen S, Kleivi K et al. Protein lysate microarray analysis to identify microRNAs regulating estrogen receptor signaling in breast cancer cell lines. Oncogene. 2009;28(44):3926-36. doi:10.1038/onc.2009.241.

30. Youl PH, Baade PD, Aitken JF, Chambers SK, Turrell G, Pyke C et al. A multilevel investigation of inequalities in clinical and psychosocial outcomes for women after breast cancer. BMC Cancer. 2011;11:415. doi:10.1186/1471-2407-11-415.

31. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65. doi:10.1038/nature11632.

32. Chacon-Cortes D, Haupt L, Lea R, Griffiths L. Comparison of genomic DNA extraction techniques from whole blood samples: a time, cost and quality evaluation study. Molecular Biology Reports. 2012;39(5):5961-6. doi:10.1007/s11033-011-1408-8.

33. Nasiri H, Forouzandeh M, Rasaee MJ, Rahbarizadeh F. Modified salting-out method: high-yield, high-quality genomic DNA extraction from whole blood using laundry detergent. Journal of Clinical Laboratory Analysis. 2005;19(6):229-32.

34. Chacon-Cortes D, Griffiths L. Methods for extracting genomic DNA from whole blood samples: current perspectives. Journal of Biorepository Science for Applied Medicine. 2014;2:1-9.

35. Huberman JA. Importance of measuring nucleic acid absorbance at 240 nm as well as at 260 and 280 nm. Biotechniques. 1995;18(4):636.

36. Sahota A, Brooks AI, Tischfield JA, King IB. Preparing DNA from blood for genotyping. CSH protocols. 2007. doi:10.1101/pdb.prot4830.

37. Sambrook J, Russell DW. Molecular Cloning: A Laboratory Manual. vol v. 1. Cold Spring Harbor Laboratory Press; 2001.

38. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W et al. An Analysis of Human MicroRNA and Disease Associations. PLoS One. 2008;3(10):e3420. doi:10.1371/journal.pone.0003420.

39. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11.

40. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics. 1992:361-72.

41. Hardy GH. Mendelian proportions in a mixed population. Science. 1908;28(706):49-50.

42. Fisher SRA, Yates F. Statistical Tables for Biological, Agricultural and Medical Research... revised and enlarged. Oliver & Boyd; 1963.

43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007;81(3):559-75. doi:10.1086/519795.

44. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263-5. doi:10.1093/bioinformatics/bth457.

Table 1 SNPs in the mir-17-92 cluster gene selected for primer design using the MassARRAY® Assay Design Suite v1.0 software (SEQUENOM Inc., San Diego, CA, USA)

miRNA	Locus	Location	miRNA SNPs in gene region	MAF	Location	SNPs near 3' end	MAF	Location	Downstream distance (bp)
hsa-mir-92a-1	13q31.3	92,003,568 - 92,003,645	rs72631821	NA	92,003,588				
			rs9589207	0.0211	92,003,589				
miR-17-92 Cluster Gene	13q31.3	92,000,074 - 92,006,829	rs4284505	0.4867	92,001,472	rs138514639	0.0517	91,998,497	1577
			rs72640333	0.0939	92,004,927				
			rs113242099	0.0504	92,005,118				
			rs111371822	0.1484	92,005,134	rs1888138	0.0916	91,999,496	578
			rs7336610	0.4524	92,005,137				
			rs7318578	0.3594	92,005,469				
			rs17735387	0.0829	92,006,054	rs2351704	0.3539	91,999,716	358
			rs1428	0.463	92,006,770				

Table 2 Primer sequences for mir-17-92 SNPs included in genotyping study using multiplex PCR reaction and MALDI-TOF MS

SNP	Forward primer sequence	Reverse primer sequence	iPLEX® (Extension) primer sequence
rs1888138	ACGTTGGATGGTTTATTACTTTACCGGCCC	ACGTTGGATGTACTTCTCTGGTTCCGGTTG	TGGGTTTCCGAGGTA
rs7336610	ACGTTGGATGAAAAAGTTCCGGCTGGACAC	ACGTTGGATGACAGCGTTTCACCATGTCGG	GACTGACCTCAGGTAATCC
rs9589207	ACGTTGGATGACTCAAACCCCTTTCTACAC	ACGTTGGATGGGACAAGTGCAATACCATAC	AAGGAACACAGCATTGCAAC
rs17735387	ACGTTGGATGAGCCTTAACTATTTGGAGGG	ACGTTGGATGGCTTTCTTTCCAAATATAGGC	GGGGAGAAAGTTGTACATGCAAA
rs4284505	ACGTTGGATGCTTTGCAGTCTCGGGTGTTC	ACGTTGGATGTGATATTGCAACGACGAGCC	TGATCCTGCCTTTTTCAGTTCCTT
rs1428	ACGTTGGATGTCAATATTCTCGTTCTGGAC	ACGTTGGATGACAGTTTGGTCTGGCTGTTT	GCATTTAATGTTAATAAATAAAATACTG

Table 3 Chromosomal location and allele information for selected miRNA SNPs in the mir-17-92 Cluster gene

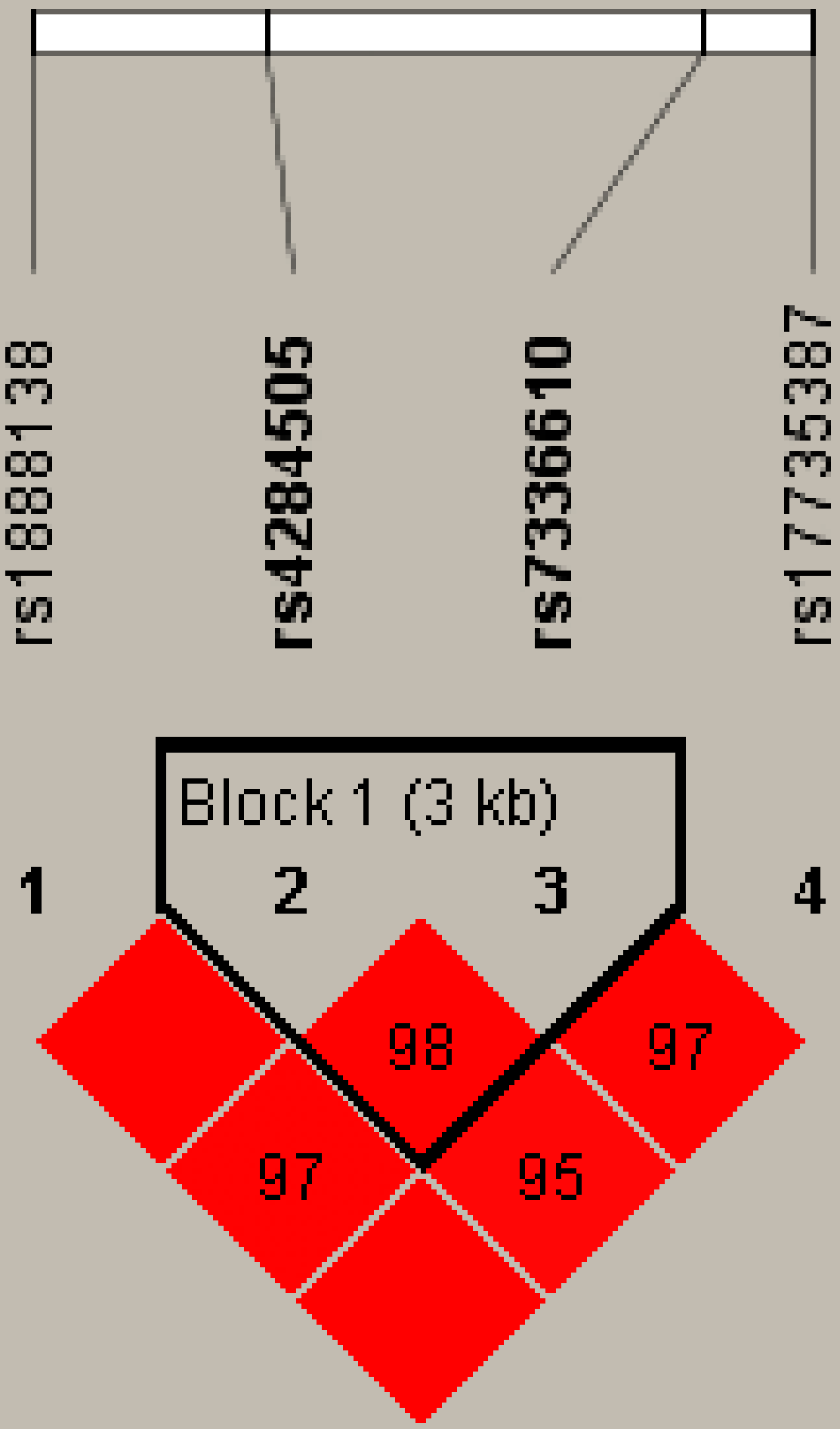
SNP	Locus	Wild Type allele	Mutant allele	Chromosomal position	MAF
rs1888138	13q31.3	A	T	chr13: 91999496	0.0916
rs4284505	13q31.3	G	A	chr13: 92001472	0.4867
rs9589207	13q31.3	G	A	chr13: 92003589	0.0211
rs7336610	13q31.3	T	C	chr13: 92005137	0.4524
rs17735387	13q31.3	G	A	chr13: 92006054	0.0829
rs1428	13q31.3	T	G	chr13: 92006770	0.4625

Table 4 Allele frequencies of SNPs in the mir-17-92 cluster gene obtained from the GRC-BC and GU-CCQ BB cohorts

		rs1888138			rs4284505			rs7336610			rs17735387			rs1428		
Population	**Alleles**	A (%)	T (%)	*p*-value	G (%)	A (%)	*p*-value	T (%)	C (%)	*p*-value	G (%)	A (%)	*p*-value	T (%)	G (%)	*p*-value
GRC-BC Population	**Control**	361 (96.5)	13 (3.5)	**0.86**	189 (53.1)	167 (46.9)	**0.01**	185 (52.6)	167 (47.4)	**0.03**	360 (96.3)	14 (3.7)	**0.68**	182 (56.2)	142 (43.8)	**0.67**
	Cases	470 (96.3)	18 (3.7)		263 (61.7)	163 (38.3)		263 (60.0)	175 (40.0)		449 (96.8)	15 (3.2)		278 (57.7)	204 (42.3)	
GU-CCQ BB Population	**Control**	590 (95.8)	26 (4.2)	**0.87**	362 (58.8)	254 (41.2)	**0.03**	346 (56.2)	270 (43.8)	**0.10**	590 (95.8)	26 (4.2)	**0.66**	383 (62.8)	227 (37.2)	**0.36**
	Cases	1295 (95.9)	55 (4.1)		767 (64.1)	429 (35.9)		712 (60.1)	472 (39.9)		1287 (96.2)	51 (3.8)		655 (60.5)	427 (39.5)	
Hapmap (%)		95.9	4.1		57.9	42.1		55.7	44.3		98.2	1.8		65.6	34.4	

Table 5 Haplotype block association analysis for selected SNPS in the mir-17-92cluster gene in combined GRC-BC/GU-CCQ BB population

	GT (%)	AC (%)	GC (%)
Control	269 (54.3)	214 (43.3)	12 (2.4)
Cases	544 (58.9)	336 (36.4)	43 (4.7)
x^2	5.05	12.21	6.58
***p*-value**	0.02	5.0E-04	0.01
Global Statistics	$x^2 = 9.21$ *p*-value = *0.01*		


rs1888138
rs4284505
rs7336610
rs17735387
Block 1 (3 kb)
1
2
3
4
98
97
97
95